Motor Trends: Automatic or a Stick? It doesn't matter.

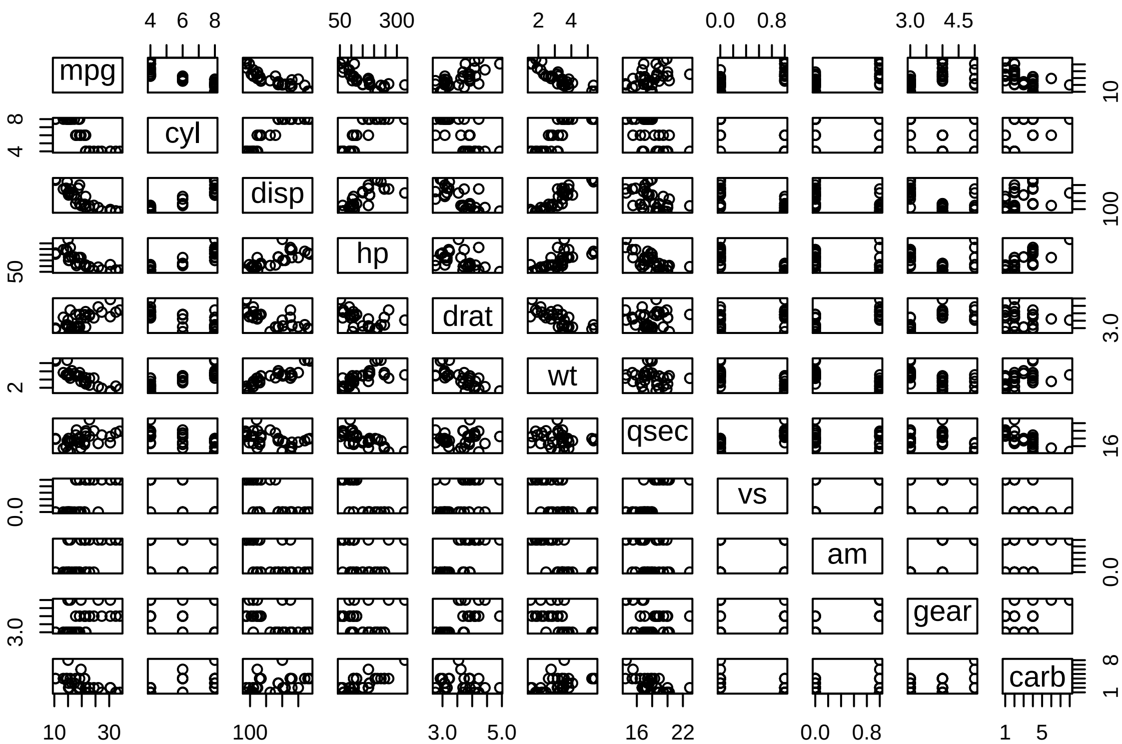
Summary

We take a look at some historical data (the 1974 Motor Trend US magazine car road tests) and try to answer what car characterstics influence fuel confumption (measured in Miles per Gallon). We will take a closer look on the relation between transmission type (auto or stick) and car's MPG and observe that if we include other important variables like car's weight and number of cylinders in car's engine, then the relation between transmission type and fuel consumption is insignificant.

Exploratory Analysis

First, we take a look at the data to see if there are some obvious correlations.

```r
data(mtcars)
pairs(mtcars)
```



Let's see correlation table for completeness:

```r
cor(mtcars)
```

```
##          mpg     cyl    disp      hp     drat      wt    qsec      vs
## mpg   1.0000 -0.8522 -0.8476 -0.7762  0.68117 -0.8677  0.4187  0.6640
## cyl  -0.8522  1.0000  0.9020  0.8324 -0.69994  0.7825 -0.5912 -0.8108
## disp -0.8476  0.9020  1.0000  0.7909 -0.71021  0.8880 -0.4337 -0.7104
## hp   -0.7762  0.8324  0.7909  1.0000 -0.44876  0.6587 -0.7082 -0.7231
## drat  0.6812 -0.6999 -0.7102 -0.4488  1.00000 -0.7124  0.0912  0.4403
## wt   -0.8677  0.7825  0.8880  0.6587 -0.71244  1.0000 -0.1747 -0.5549
```

```
## qsec  0.4187 -0.5912 -0.4337 -0.7082  0.09120 -0.1747  1.0000  0.7445
## vs    0.6640 -0.8108 -0.7104 -0.7231  0.44028 -0.5549  0.7445  1.0000
## am    0.5998 -0.5226 -0.5912 -0.2432  0.71271 -0.6925 -0.2299  0.1683
## gear  0.4803 -0.4927 -0.5556 -0.1257  0.69961 -0.5833 -0.2127  0.2060
## carb -0.5509  0.5270  0.3950  0.7498 -0.09079  0.4276 -0.6562 -0.5696
##            am    gear     carb
## mpg   0.59983  0.4803 -0.55093
## cyl  -0.52261 -0.4927  0.52699
## disp -0.59123 -0.5556  0.39498
## hp   -0.24320 -0.1257  0.74981
## drat  0.71271  0.6996 -0.09079
## wt   -0.69250 -0.5833  0.42761
## qsec -0.22986 -0.2127 -0.65625
## vs    0.16835  0.2060 -0.56961
## am    1.00000  0.7941  0.05753
## gear  0.79406  1.0000  0.27407
## carb  0.05753  0.2741  1.00000
```

MPG seems to be highly correlated with number of cylinders (cyl), displacement (disp), horsepower (hp) and weight (wt), so we will use those variables in our regression models. Transmission type doesn't seem to be highly correlated, but we will include it in our model to be able to answer our research question.

Models

We will create a series of regression models of increasing complexity. We will start with the simplest one, where the MPG is being explained only by the transmission type.

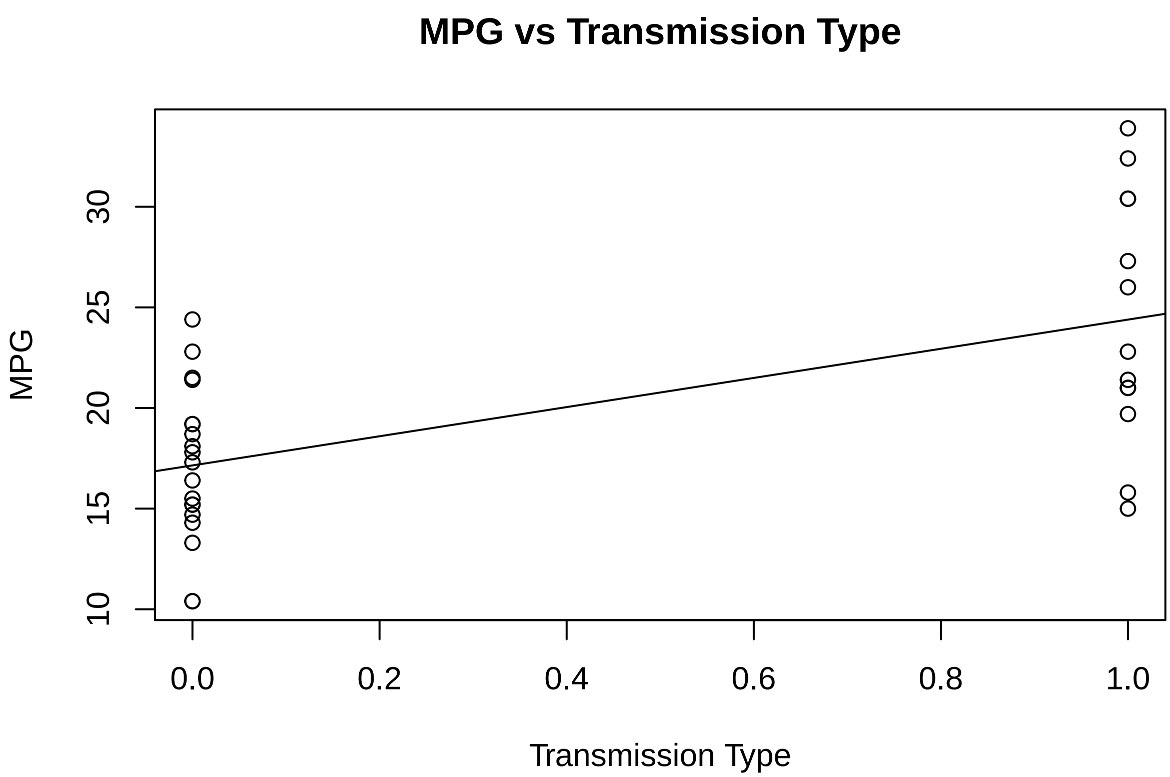
```r
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## am              7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

This simple model is highly interpretable - intercept coeficient is 17.1474, which means that a car with automatic transmission (am == 0) is expected to drive 17.1474 miles per gallon and am coefficient is 7.2449, which means that a car with manual transmission (am == 1) is expected to drive 7.2449 more miles per gallon. All p-values are much lower than 0.05, so the model seems to be strong, but the variation explained (R-squared) is only 0.3598, so we need to include more variables to have a more complete view on cars' MPG.

Let's see the relation beetween transmission type and MPG on a plot.

```
plot(mtcars$am, mtcars$mpg, xlab="Transmission Type", ylab="MPG", main="MPG vs Transmission Type")
abline(fit)
```



We will try to create a series of more complex models to get a better view on fuel consumption. We will regress on variables centered on mean or minimal value of those variables, so the intercept would remain interpretable - it doesn't have sense to have an intercept that is a MPG value for car with zero horsepower, zero weight and zero cylinders, but it does have sense to have an intercept as a value for car with mean horsepower (146.6875 HP) and weight (3217.25 lb) and minimal number of cylinders (4 cylinders).

```
mtcars$cyl.2 <- mtcars$cyl - min(mtcars$cyl)
mtcars$disp.2 <- mtcars$disp - mean(mtcars$disp)
mtcars$hp.2 <- mtcars$hp - mean(mtcars$hp)
mtcars$wt.2 <- (mtcars$wt - mean(mtcars$wt)) * 1000 # also scale to lb
```

We add variables to regressin model in order of correlation and test their significance using ANOVA test:

```
fit2 <- lm(mpg ~ am + wt.2, data = mtcars)
fit3 <- lm(mpg ~ am + wt.2 + cyl.2, data = mtcars)
fit4 <- lm(mpg ~ am + wt.2 + cyl.2 + disp.2, data = mtcars)
fit5 <- lm(mpg ~ am + wt.2 + cyl.2 + disp.2 + hp.2, data = mtcars)
anova(fit, fit2, fit3, fit4, fit5)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt.2
## Model 3: mpg ~ am + wt.2 + cyl.2
## Model 4: mpg ~ am + wt.2 + cyl.2 + disp.2
## Model 5: mpg ~ am + wt.2 + cyl.2 + disp.2 + hp.2
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     29 278  1       443 70.54   7e-09 ***
## 3     28 191  1        87 13.91 0.00094 ***
## 4     27 188  1         3  0.42 0.52370
## 5     26 163  1        25  4.03 0.05510 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we can see, adding weight and cylinder information results with significanlty better models (p-value $<$ 0.05) and adding displacement and horse power does not. In result, we use fit3 model using transmission type, weight and number of cylinders as our final model.

Let's see it's details:

```r
summary(fit3)
```

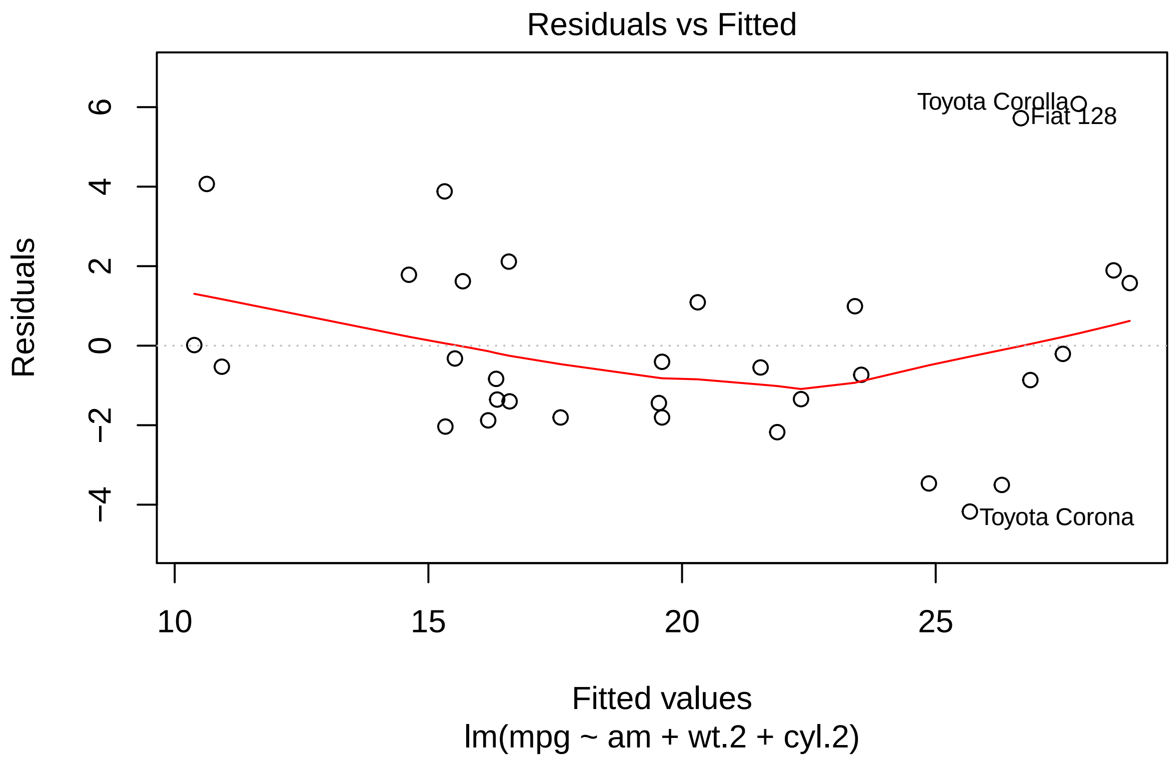
```
##
## Call:
## lm(formula = mpg ~ am + wt.2 + cyl.2, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.173 -1.534 -0.539  1.586  6.081
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.322587   1.142517   20.41   <2e-16 ***
## am           0.176493   1.304451    0.14   0.8933
## wt.2        -0.003125   0.000911   -3.43   0.0019 **
## cyl.2       -1.510246   0.422279   -3.58   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.61 on 28 degrees of freedom
## Multiple R-squared:  0.83,   Adjusted R-squared:  0.812
## F-statistic: 45.7 on 3 and 28 DF,  p-value: 6.51e-11
```

The intercept intercept coeficient is 17.1474, which means that a car with automatic transmission (am == 0), 4 cylinder engine and average weight is expected to have fuel efficiency of 23.3226 miles per gallon.

The transmission type coefficient is only 0.1765 and in the first model it was 17.1474, which means that the newly added variables (weight and number of cylinders were confounding the transmission type variable). The p-value of this coefficient is very high (0.89334), so it means that after taking an account of car's weight and number of cylinders the relation between MPG and transmission type diminishes.

The final model seems to be strong with R-squared value of 'r summary(fit3)$r.squared' and resudual plot without anomalies:

```r
plot(fit3, which = 1)
```



Conclusion

Is an automatic or manual transmission better for MPG? On the surface, there is a difference, but if you look in depth and take an account for car's weight and engine's number of cylinders, the transmission type of your car doesn't matter.